Available online at www.sciencedirect.com

ScienceDirect

(2015) –13

Journal Logo

Characterizing ordering in liquids: an information theoretic approach

Luis Carlos Pardo

e mail: luis.carlos.pardo@upc.edu

Grup de Caracterització de Materials, Departament de Física i Enginyeria Nuclear, ETSEIB, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Catalonia, Spain

Andrés Henao

Grup de Simulació per Ordinador en Matèria Condensada, Departament de Física i Enginyeria Nuclear, B4-B5 Campus Nord, Universitat Politècnica de Catalunya, E-08034 Barcelona, Catalonia, Spain

Alessandro Vispa

Grup de Caracterització de Materials, Departament de Física i Enginyeria Nuclear, ETSEIB, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Catalonia, Spain

Abstract

The determination of special molecular arrangements in disordered phases such as liquids is inherently difficult due to its lack of periodicity, in contrast to the crystalline solids. We have already settled a general method to study molecular liquids capable to unveil the details of the molecular ordering from small molecules to systems as big as a protein. However it would be desirable to extract some general features of a liquid phase without going into such details. In this work we propose a method to achieve this challenge by analyzing the probability distributions describing position and orientational molecular ordering within the framework of information theory.

Keywords: 61.20.-p, 61.25.Em, 02.70.Ns, 02.60.-x Short range order, structure, liquid, information theory

1. Introduction

Solid crystalline phases are composed by molecules whose centers of mass are forming a long range ordered lattice being their relative molecular orientations fixed. For this phase the details of molecular interactions are important and determine the solid structure. On the contrary, no precise knowledge about molecular interaction is necessary to describe the structure of gases: it is basically ruled by the shape of the molecule. Liquid structure is between these two well known phases. Moreover if a material in a liquid phase is cooled down fast enough it can undergo through a glass transition where at a human time scale molecules seem not to be moving. There is a lack of a universal theory that describes liquids and how they fall out of equilibrium forming a glass, but what seems to be sure is that liquid structure is suspected to play an important role [1]. In any case there is not a unique way to characterize the structure of a liquid and many different approaches have been used in the past [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. Among molecular liquids, carbon tetrachloride occupies a special place: it was the first one to be studied by means of an incipient X-ray diffraction technique. A great amount of works have been published on the relative molecular ordering of CCl_4 at distances close to a central molecule [16, 17, 18, 19, 20, 21, 22]. All these works have in common that the obtained structure

respects the tetrahedral symmetry of the molecule, being the closest neighbor located either in the corners of the tetrahedron defining the molecule [17, 19], in the edges [21] or in front of its faces [18, 20]. Although it is crucial to know the detailed molecular ordering in a liquid, in this work we would like to develop a frame to study molecular liquids without entering in its detail so that we can capture its general features. We will do that by using an information theoretic approach to study the six-dimensional probability distribution function (PDF) describing the relative position and orientation of a molecule at a distance r from a central one $g\left(r, \Omega_{pos}, \Omega_{or}\right)$, where Ω_{pos} are the angles θ_{CM} and ϕ_{CM} describing the position of the center of mass of a neighbor molecule with respect to the central one at a distance r and Ω_{or} are the three Euler angles θ_{or}, ϕ_{or} and ψ_{or} describing the relative molecular orientation.

To do that, the first step would be to obtain the molecular configurations that are describing microscopically the system. Such configurations can be obtained from Molecular Dynamics simulations using parameters that produce macroscopic properties compatible with the experiment or using reverse methods such as Reverse Montecarlo (RMC) [23] or an Empirical Potential Structure Refinement [24]. We would like to point out that although we have chosen carbon tetrachloride as an example to perform our analysis this method is a general tool for describing the liquid structure.

2. The data: Molecular dynamics simulation

In order to obtain the molecular configurations necessary to calculate $g\left(r, \Omega_{pos}, \Omega_{or}\right)$ we performed an MD simulation using the Gromacs 4.5 [25] package. The potential parameters used were for Carbon σ = 3.7746Å, ϵ = 0.2271kJ/mol, q = −0.696e and for Chloride: σ = 3.4667Å, ϵ = 1.0944kJ/mol and q = 0.174e. The simulation was made on a 216 rigid molecules system using the NPT ensemble at the thermodynamics conditions of the liquid, namely T = 298K and P = 1atm.
The chosen time step to perform the simulation was Δt = 5fs (during 1200ps), a test run with a smaller time step of Δt = 1fs has been done remaining the results of this work unchanged. We used a switched cut-off from 8Å to 14Å for Lennard-Jones interactions and 14Å for coulomb pairs. We used the Particle Mesh Ewald (PME) method beyond the electrostatic cut-off for the reciprocal space sum.

3. Detailed determination of molecular ordering

Contrary to what happens in a solid crystalline phase where it is necessary only to study the unit cell to reproduce the long range structure, in liquids it is necessary to study the whole system to get some information about the most probable arranging of molecules, and this must be done by using probability distributions. Although it is not the goal of this work, we will now shortly describe a general method to study the details of molecular arranging in a liquid for the sake of clarity of the later discussion. This general method is suitable to deal with small molecules such as Carbon Tetrachloride as well as with molecules as large as proteins [26] (more details can be found in our previous works [15, 27, 28]). For doing that it is necessary to attach an axis set to each molecule. In figure 1 we show the axis chosen in our case: Z axis is set in the bisecting angle of two C-Cl bonds, Y axis is set perpendicular to this axis and coplanar with a Cl-C-Cl plane, and X axis is determined as usually in an orthonormal axis set ($X = Y \times Z$). We then analyze the position and orientation of a neighbour molecule setting each molecule as reference in all configurations (in our case 1000), i.e. we calculate $g\left(r, \Omega_{pos}, \Omega_{or}\right)$. However instead of choosing r to characterize the distance from the central molecule we choose the Molecular Coordination Number (MCN). In order to calculate the MCN we have ordered neighbour molecules by their distance to the central molecule and then we have numbered them to calculate it. This has two main advantages: first it eliminates trivial effects of density when comparing a liquid at different temperatures [28]. On the other hand, the structure of a liquid changes fast at short distances and slower for molecules far away from the central molecule. Choosing the MCN eliminates those effects expanding the short distance region and shrinking the long distance region. We therefore calculate $g\left(MCN, \Omega_{pos}, \Omega_{or}\right)$ for $MCN = 1, 2, \cdots, n$ so that we have a five-dimensional probability distribution describing the position and orientation as function of MCN. In panel (a1) of figure 1 we show the two-dimensional PDF $g\left(\cos(\theta_{CM}), \phi_{CM}\right)$ describing the molecular position. We show in the same figure in panel (a2) how high probability regions of $g\left(\cos(\theta_{CM}), \phi_{CM}\right)$ are related to the position of the center of mass of neighboring molecules.

In order to study the molecular orientation we must choose the highest probability regions of $g\left(\cos(\theta_{CM}), \phi_{CM}\right)$, i.e. some specific molecular positions. In panels (b1) and (b2) of figure 1 we show

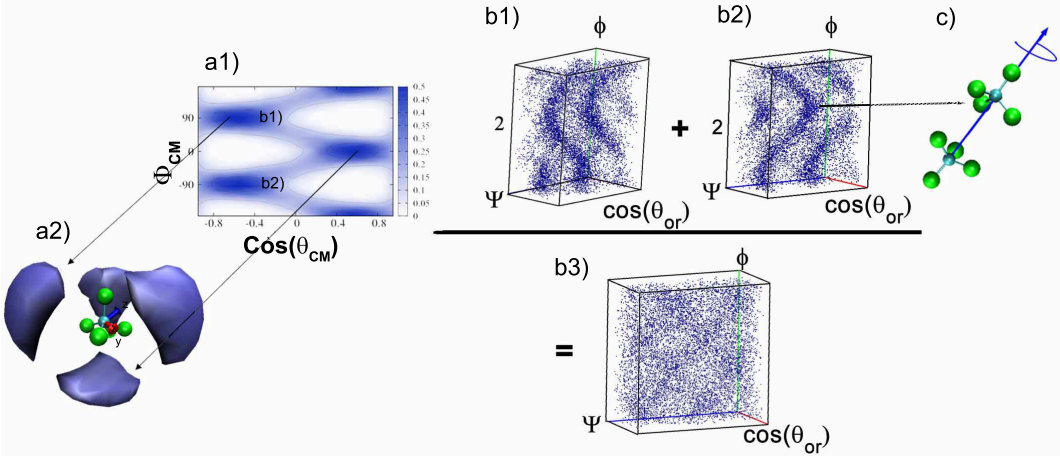

Figure 1. (Color online) (a1) Probability distribution of the angles that determine molecular position $\cos(\theta_{CM})$ and ϕ_{CM} for the closest four molecules to a central one. In figure (a2) we show how this probability map yield to the 3D structure of the liquid at short distances. Figures (b1) and (b2) show the 3D orientational probability distributions of the angles determining molecular orientation θ_{or}, ϕ_{or} and ψ_{or} for different regions of the positional map shown in figure (a1). In panel (b3) we show the merging of all orientational probability distributions $g(\Omega_{or})$ used for the calculations of entropy and mutual information.

the 3D probability distribution for the Euler angles describing the molecular orientation for molecules in the highest probability positions labeled as (b1) and (b2) in $g(\cos(\theta_{CM}), \phi_{CM})$. Without entering into details (see [22]) the obtained spirals represent a series of probable orientations that correspond to a molecule rotating around its C-Cl axis with its face parallel to the central molecule face. In figure (b3) we show the complete 3D probability distribution, thus not selecting a certain molecular position. As it can readily be seen, it is impossible from this PDF to extract any detailed feature concerning molecular orientations, but as we will see, some general information can be obtained from such a 3D map about the liquid structure.

4. Information theory

Information theory is considered to be born with the seminal paper of Claude E. Shannon in the early 50's [29]. This theory was aimed to quantify the amount of information that a message carries and how can it be transmitted through a noisy channel without a significant loss of quality. However this theory quickly crossed its borders and it is now applied to a vast number of fields such as message encryption [30], analysis of seismic data to search oil [31], and to decide which is the best strategy to follow in gambling games [32]. In the context of the present work, information theory will be used to quantify both the amount of information carried by a n-dimensional probability distribution and the correlation between two or more variables that characterize the short range order of a liquid. In this section we will recall some basic notions of information theory following the excellent paper of Matsuda et al. [33].

4.1. Entropy

Let's assume that we have a n-dimensional PDF described by variables $A_i, i = 1, \ldots, n$ and that each variable A_i can take the discrete values a_i being the number of these values for each variable a_i equal to $\tilde{N}_i$. In this case the entropy associated to the normalized discrete probability distribution $p(a_1, \ldots, a_n)$ will be:

$$S(A_i) = -\sum_{\{a_i\}} p(a_i) \ln p(a_i). \tag{1}$$

The entropy so defined has an upper and a lower bound depending on the way the PDF has been generated. To study them let's assume that the number of total bins of the n-dimensional distribution is N (so that $N = \sum_{i=1}^{n} \tilde{N}_i$). If the probability distribution is completely flat, i.e. if all the values of probability are $p(a_1, \ldots, a_n) = 1/N$, the PDF contains no information and in this case the associated entropy will be maximal and equal to lnN:

$$S(A_1 \ldots A_n) = -\sum_{k=1}^{N} \frac{1}{N} \ln \frac{1}{N} = \ln N. \tag{2}$$

On the other hand if we have a probability distribution with only one bin, thus with probability one, and

all the rest with null probability, the entropy associated to the PDF will be zero. Therefore we can also regard the entropy as a measurement of how "peaky" the landscape of a PDF is: smooth probability distributions will have high entropies. This fact is used, for example, in maximum entropy methods to find the smoothest probability distribution able to describe a data set.

An analogous definition can be used to calculate the entropy of any subset of variables $A_{1,\dots,m}$ where $m < n$, just marginalizing the probability distribution by summing the variables $A_{m+1,\dots,n}$ and defining thus the entropy of this reduced ensemble as:

$$p_{a_1,\dots,m} = \sum_{\{a_{m+1},\dots,n\}} p\,(a_{m+1},\dots,n)\,. \tag{3}$$

In this case the limits for the entropy hold, being in this case N the number of bins of the m-dimensional probability distribution.

4.2. Mutual information

Mutual information quantifies the correlation among several variables whose dependence is encoded in a certain PDF. For the sake of clarity we will begin describing the two-dimensional case, then we will extend the description to higher dimensions. For 2D probability distributions the mutual information is defined as:

$$\begin{aligned} I_2(A_1,A_2) &= \sum_{a_1,a_2} p(a_1,a_2)\, ln \frac{p(a_1,a_2)}{p(a_1)p(a_2)} \\ &= S(A_1) + S(A_2) - S(A_1A_2). \end{aligned} \tag{4}$$

If variables A_1 and A_2 are independent, we can write their joint probability as $p(a_1,a_2) = p(a_1)\cdot p(a_2)$ and in this case the calculated mutual information is zero. On the other hand, the maximum value is reached when the variables are fully correlated, and such a value is the smallest between the individual entropies of variables A_1 and A_2, i.e:

$$0 \le I_2(A_1,A_2) \le min\{S(A_1), S(A_2)\} \tag{5}$$

For the three-dimensional case, the mutual information for a probability distribution is calculated taking into account its relationships with the entropies of the three-dimensional PDF, and its projections in two and one dimensions [33]:

$$\begin{aligned} I_3(A_1,A_2,A_3) = {} & S(A_1) + S(A_2) + S(A_3) \\ & - S(A_1A_2) - S(A_1A_3) - S(A_2A_3) \\ & + S\,(A_1A_2A_3)\,. \end{aligned} \tag{6}$$

Again, the mutual information associated with a three-dimensional probability distribution measures how well correlated are the three variables. However, in this case, the limits for the mutual information are very different than those of the two-dimensional case since the lower bound can be negative [33]:

$$\begin{aligned} -min\{S(A_1), S(A_2), S(A_3)\} & \\ \le I_3(A_1,A_2,A_3) \le & \\ min\{S(A_1), S(A_2), S(A_3)\}\,. & \end{aligned} \tag{7}$$

A negative value of mutual information in an N-dimensional probability distribution is associated with the frustration of the variables describing the PDF. If the probability distribution of the quantities describing the orientation and position of two molecules would have a negative value of mutual information, it would mean (following [33]) that the variables describing the geometry of the relative position and orientation of two molecules are frustrated. This must be understood as it is stated in the Appendix of this work, i.e. that knowing two variables the third remains undetermined.

Mutual information also allows us to write the entropy of an n-dimensional PDF as the sum of the entropies of the one-dimensional projection of the PDF plus mutual information terms:

$$\begin{aligned} S_{tot} = {} & \sum_{i=1}^{n} S_i(A_i) - \sum_{i<j} I_2(A_i,A_j) \\ & + \sum_{i<j<k} I_3(A_i,A_j,A_k) + \cdots \end{aligned} \tag{8}$$

This relationship allows us to calculate the entropy of a high-dimensional PDF using lower dimensional PDFs (being sometimes *high-dimensional* as low as $n \ge 4$). This is interesting because the number of n-dimensional voxels increases as N^n which leads to two problems. First, since the memory of the computer and its speed is finite it will not be easy to handle n-dimensional matrices if we want to have a reasonable number of voxels N. On the other hand calculations of PDFs are usually done by counting events and then normalizing by their total amount. The number of events we must count to have a reasonable PDF increases with its dimensionality (in fact it increases as N^n). If the number of events is not high enough the PDF will consist in series of sparse switched on voxels making any calculation of entropy or mutual information meaningless.

In order to make clear the physical meaning of the quantities entropy and mutual information we have included in this work an appendix where these quantities

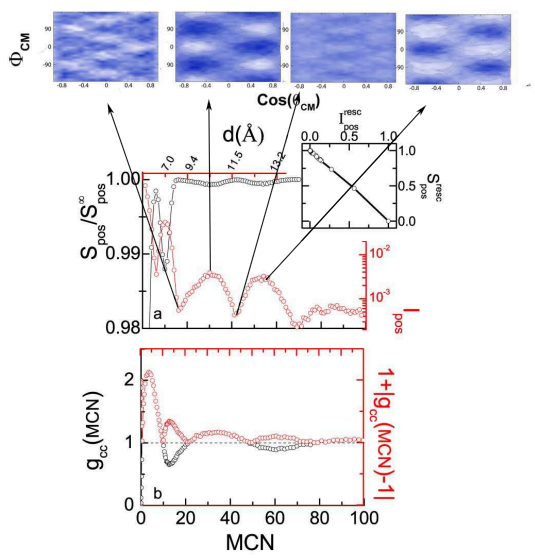

Figure 2. (Color online) In panel (a) we show both the rescaled entropy and the mutual information for the 2D probability distributions describing the molecular position such as that of panel (a1) of figure 1 as a function of the molecular coordination number (MCN). The inset shows the rescaled mutual information as a function of the rescaled entropy for all studied distances (see text for the rescaling definition in this case). In panel (b) we show the radial distribution function $g_{CC}(MCN)$ together with its flipped version $\|g_{CC}-1\|+1$. The insets show the positional maps $g\,(\cos(\theta_{CM}),\phi_{CM})$ for maxima and minima of the mutual information function.

are calculated for some characteristic probability distributions in two and three dimensions.

5. Liquid ordering and mutual information

So far, we have established a detailed method to determine how do molecules distribute around a central one in a liquid and a way to characterize probability distributions. We are now interested in extracting some general information from the 6D probability distribution $g\left(r,\Omega_{pos},\Omega_{or}\right)$. In this section we will analyze some features of the liquid structure of CCl_4. If we restrict ourselves only to the first three terms of equation 8 we can define the positional S_{pos}, the orientational S_{or} and the mixed contribution $S_{pos\,or}$ to the total entropy as follows:

$$S_{pos} = S\,(\cos(\theta_{CM}))+S\,(\phi_{CM})-I\,(\cos(\theta_{CM}),\phi_{CM}) \quad (9)$$

$$\begin{aligned} S_{or} &= S\,(\cos(\theta_{or})) + S\,(\phi_{or}) + S\,(\Psi_{or}) \\ &\quad -I\,(\cos(\theta_{or}),\phi_{or}) - I\,(\cos(\theta_{or}),\psi_{or}) - I\,(\phi_{or},\psi_{or}) \\ &\quad I\,(\cos(\theta_{or}),\phi_{or},\psi_{or}) \end{aligned} \quad (10)$$

$$\begin{aligned} S_{pos\,or} &= S_{tot} - S_{pos} - S_{or} \\ &= -I_{pos_1,or_1} + I_{pos_1,or_2} + I_{pos_2,or_2} \end{aligned} \quad (11)$$

where I_{pos_1,or_1}, I_{pos_1,or_2} and I_{pos_2,or_1} are mutual information terms having 1, 1 or 2 positional variables and 1, 2 or 1 orientational variables, respectively. All contributions to $S_{pos\,or}$ are thus coming exclusively from mutual information terms of order two and three. We would like to point out that the calculated entropy from $g\left(r,\Omega_{pos},\Omega_{or}\right)$ is related to the thermodynamic excess entropy (i.e. that deviating from the gas phase) if we restrict ourselves to two-body molecular interactions [14].

We will first start analyzing the positional contribution to the total entropy. To do that we show in the panel (a) of figure 2 the entropy scaled to its asymptotic value for long distances (S_{pos}/S^{∞}_{pos}), and the mutual information of the PDF related to the positional variables $\cos(\theta_{CM})$ and ϕ_{CM}. As expected, for long distances mutual information tends to zero and entropy to a fixed and maximal value. From the same figure we also see that both mutual information and entropy are correlated, so that when one increases the other decreases.

In order to investigate this fact we show in the inset of this figure a different rescaling for both magnitudes so that they have a limited range variation from zero to one ($\xi^{resc} = \xi - \xi_{min}/\xi_{max} - \xi_{min}$, being ξ either the entropy or the mutual information). From this figure we can see that they are completely correlated so that $S^{resc} = -I^{resc}$. This is due to the small differences between the values of entropies for short and long distances coming from the 1D projections of $g(\cos(\theta_{CM}),\phi_{CM})$, i.e. $S\,(\cos(\theta_{CM}))$ and $S\,(\phi_{CM})$. In this case equation 9 can be simply rewritten as $S_{pos} = ct - I\,(\cos(\theta_{CM}),\phi_{CM})$.

We have plotted in panel (b) of the same figure the partial radial distribution of the carbon atoms of CCl_4, i.e. the center of mass of the molecule. Comparing both panels (a) and (b) we can see that for long distances minima and maxima of $g_{CC}(r)$ are correlated with maxima (minima) of the mutual information (entropy) from

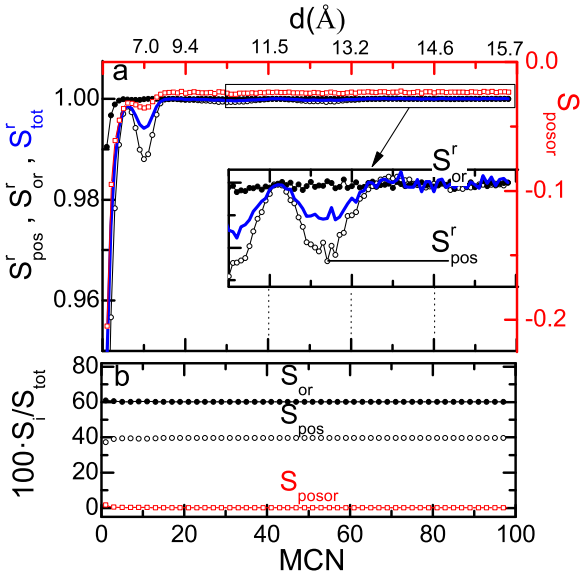

Figure 3. (Color online) (a) Orientation (filled circles), positional (empty circles) and mixed term (empty squares) contributions to the total entropy (thick line). The inset shows a zoom of the high distance region. In panel (b) we show the percentage of the contribution to the total entropy of each term (see text). The upper axis concerning distance d(Å) is for the help of the reader

the probability distribution describing the position of the molecular centers of mass. In order to help the comparison we plot in the same figure $\|g_{CC}(r) - 1\| + 1$ to flip up the minima of g_{CC}. This figures tells us two important things about this simple molecular liquid:

- The liquid does not get monotonically more disordered as the distance from a central molecule increases. On the contrary in some distance intervals the liquid does order itself.
- The regions where the liquid is more ordered (those with a maximum in mutual information or a minimum in entropy) are correlated with regions of high or low density described by maxima and minima of the radial distribution function of the center of mass.

To make clear these two points we have included as insets in figure 2 the 2D-PDF describing the position of the centers of mass $g(\cos(\theta_{CM}), \phi_{CM})$ as those in figure 1 in maxima and minima of mutual information (entropy). From these figures it is clear that minima (maxima) of mutual information (entropy) are related to regions where there is not a particular ordering of the centers of mass of neighboring molecules (we get more or less flat distributions). On the contrary around maxima (minima) of mutual information (entropy) it is clear that the liquid is more ordered.

We will now investigate how do the positional S_{pos} orientational S_{or} and the mixed term $S_{pos\,or}$ contribute to the total entropy of the system. To do that we show in panel (a) of figure 3 the three contributions and the total entropy rescaled to their asymptotic values for long distances. As it can be seen in the figure the orientational contribution to the entropy fades out much faster than the positional one, being structured only for distances smaller than circa 6Å. If we zoom the region for distances far away from the central molecule we also see no structure for the orientational contribution. Maybe, the most astonishing thing from this figure is that the crossed contribution to the entropy $S_{pos\,or}$ is negative. Therefore the contributions of mutual information from the three-dimensional projections of the probability distribution $g\left(MCN_{fixed}, \Omega_{pos}, \Omega_{ori}\right)$ must be either negative or smaller than their two-dimensional counterparts. We will come again to this point later in this section. In order to quantify the contribution of each term we have calculated the percentage to the total entropy of each term. However, since the contribution of the crossed term $S_{pos\,or}$ is negative, we do not normalize the contributions to the total entropy, but to the quantity $S_{pos} + S_{or} + \left\|S_{pos\,or}\right\|$. The contributions to the total entropy are, ordered by its importance, coming from the orientational, the positional and the crossed term. This can be rationalized by the simple fact that the dimensionality of the orientational PDF is higher than that of the positional contribution. In summary: *although positional contribution to the total entropy is less important than the orientational one, it is more structured and it is better correlated with the partial radial distribution function* $g_{CC}(r)$.

As we have seen, the contribution of the crossed term $S_{pos\,or}$ is negative, and as previously pointed out this can be related to a negative contribution of higher order mutual information terms. We have plotted in figure 4 all the three variable mutual information terms and all of them are negative for all distances. Following the discussion presented in last section this means that the 3D projections of the original 5D probability distribution $g\left(MCN_{fixed}, \Omega_{pos}, \Omega_{or}\right)$ are so complex that their projections in a 2D space do not allow to directly reconstruct the liquid structure. This can readily be seen in panel (b3) of figure 1 where we plotted $g(\cos(\theta_{or}), \phi_{or}, \psi_{or})$: it is even impossible to recognize any pattern from this figure since the 3D projections of the complete probability distribution merges the orien-

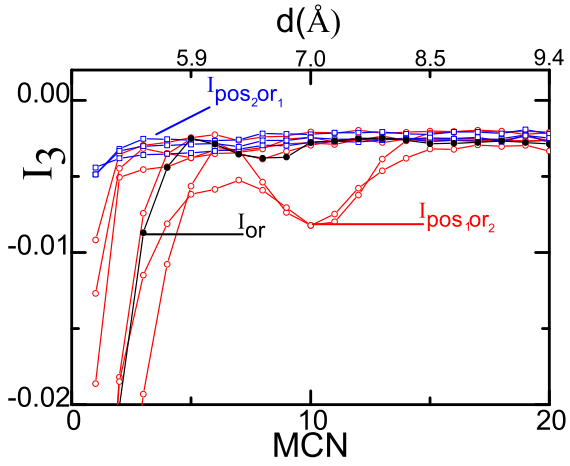

Figure 4. Three variable mutual information calculated from 3D projections of the five-dimensional probability distribution $g\left(\Omega_{pos}, \Omega_{or}\right)$. We have distinguished among 3D probability distributions fully orientational (I_{or}) and those with two or three orientational variables ($I_{or_2 pos_1}$ and $I_{or_1 pos_2}$ respectively).

tation of molecules located in different positions, or in other words, located in different regions of the 5D PDF. We can therefore state that *the structure of a liquid is geometrically frustrated*. The idea that somehow the liquid is a frustrated system is not new, and can be found, for example in [34] associating frustration to the lack of ability of the liquid to tessellate space or in [35] for liquids in a curved space, but in our case we quantify the frustration by means of the value of mutual information of three variables probability distributions. In figure 4 we see that the terms containing two or three orientational variables (θ_{or}, ϕ_{or} or ψ_{or}) have a more important contribution to the total entropy than those containing only one orientation variable. Therefore the higher contribution to the frustration of the liquid structure is the one coming from the orientation in the current case.

6. Conclusions

In this work we propose a new methodology to study the ordering of liquids up to rather long distances: the use of information theory to study the position and orientation of molecular liquids encoded in the six-dimensional probability distribution $g\left(r, \Omega_{pos}, \Omega_{or}\right)$. We have analyzed under this new procedure the structure of the first studied molecular liquid, carbon tetrachloride, arriving to two main conclusions:

- The positional structure of the studied liquid is not simply more disordered as we go far away from a central molecule, but it has regions where its center of mass are more ordered, and they are correlated with minima and maxima of the radial distribution function of their centers of mass.
- Liquids are geometrically frustrated in the sense that mutual information contributions calculated from the three-dimensional projections of $g\left(r, \Omega_{pos}, \Omega_{or}\right)$ are negative. Moreover frustration is mainly coming from orientational degrees of freedom.

7. Appendix

7.1. *Some examples of calculations of entropy and mutual information in selected probability distributions*

We add in this appendix some examples of calculations of mutual information and entropies in some 2D and 3D distribution functions to make clear their physical meaning.

7.2. *The two-dimensional case*

In figures 5 and 6 we show two cartoons of two-dimensional probability distributions ordered by increasing values of entropy $S(A_1A_2)$ and mutual information $I(A_1, A_2)$. It should be pointed out that the values of entropy and mutual information do not determine univocally the shape of a PDF. Thus, the figures are only intended to show how the shape of certain simple PDFs affects both their associated entropy and mutual information. To keep things as simple as possible, we add another restriction: the probability distributions will be generated by allowing their pixels to be only switched on or off, in other words we do not allow pixels in "grey". That means that the calculations of the entropies are straightforward for both one and two-dimensional normalized PDF's: $S(A_1) = S(A_2) = S(A_1A_2) = N_{on}$ where N_{on} are the number of pixels switched on. If we work only with a 2-dimensional PDF with the same number of pixels per side N, the maximum number of switched on pixels for $p(a_1, a_2)$ will be N^2 and thus $S_{max}(A_1A_2) = 2 \ln N$ and $S_{max}(A_1) = S_{max}(A_2) = \ln N$. Therefore the maximum mutual information possible for such a PDF is simply $I_2^{max}(A_1, A_2) = \ln N$.

We begin first with the simplest case where the variables of the probability distribution are not correlated, i.e., $I(A_1, A_2) = 0$ (see figure 5). First panel (a) shows the only PDF with $S(A_1A_2) = I(A_1, A_2) = 0$: a probability distribution with a unique pixel switched on. For this case all entropies are zero, and so it is the mutual

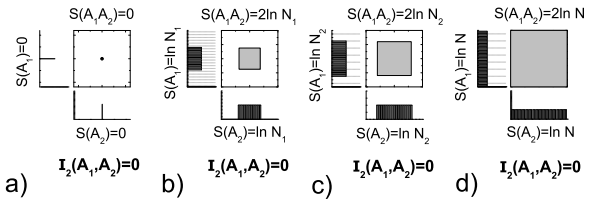

Figure 5. Two variable mutual information and entropy of some selected two-dimensional probability distributions together with their one dimensional projections. The distribution were chosen to keep zero mutual information and an increasing value of entropy. In the figure we include the calculations of the entropies for the 2D probability distributions together with the calculations for the 1D projections and the mutual information calculations for all the cases following equation 4

information. One way to increase the entropy keeping the mutual information equal to zero is simply to switch on pixels keeping the squared shape of the PDF (and, in fact, keeping any other symmetric shape) until all pixels are on, being thus the entropy maximal.

In figure 6 we show another way to increase the entropy keeping mutual information equal to zero is to switch on pixels forming a line perpendicular to any of the two variable A_1 or A_2. In this case zero mutual information reflects the fact that the knowledge of the value of one variable does not help at all to determine the value of the other one. In order to increase the entropy keeping the mutual information equal to zero one can make such a line "thicker" by switching on adjacent pixels.

On the other hand, if we want to increase the mutual information by keeping the entropy constant we can simply increase the slope of the horizontal line of figure 6b: the number of pixels switched on does not change and therefore $S(A_1A_2)$ remains constant and equal to $\ln N$. However the number of bins for the projected PDF's A_1 and A_2 start to grow and so it does their associated entropies. Maximum mutual information is reached when the line has a slope equal to one 6c: in that case knowing A_1 completely determines the knowledge of A_2. It should be pointed out that the definition of mutual information is also capable to handle with the case when two variables are not linearly correlated. In this case (see the dashed line of figure 6c) mutual information keeps being maximum. Therefore in order to discover correlations between variables it is better to use the mutual information than the correlation function $\sigma_{A_1A_2} = \sum_{a_1,a_2} (\overline{a_1} - a_1) \cdot (\overline{a_2} - a_2)$.

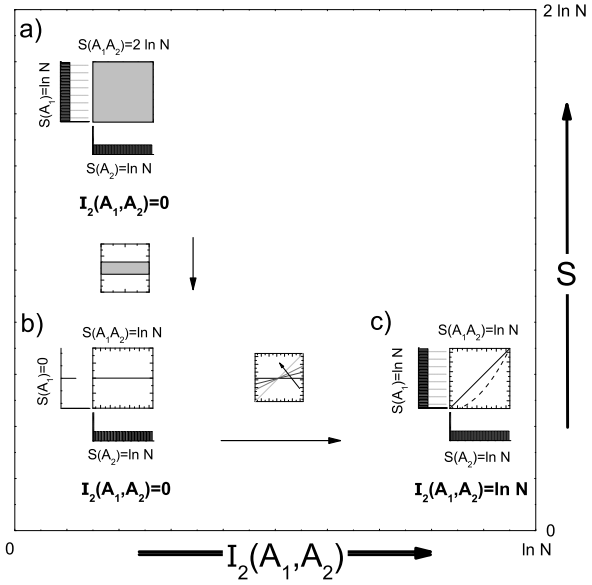

Figure 6. Examples of two-dimensional probability maps ordered by increasing values of entropy (ordinate) and mutual information (abscissa). As in figure 5 we include both 2D probability distributions together with their 1D projections and the calculations of their entropies. The inset between panels (b) and (c) indicates a 2D a series of 2D distributions with increasing mutual information in the sense of the arrow. In panel (c) we show two 2D distributions with a maximal mutual information: one with a linear correlation between variables (full line) and one with a non-linear correlation (dashed line).

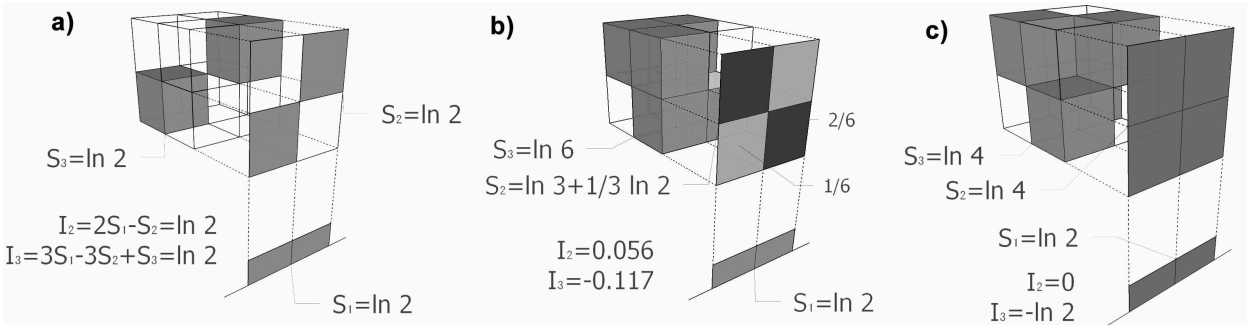

Figure 7. Symmetric three-dimensional probability distributions together with their two and one-dimensional projections, ordered by decreasing values of mutual information. Figures a and b are coming from a three spin system at zero temperature in the non-frustrated and frustrated case respectively. Figure c is one example of the maximum frustration that can be achieved from a three-dimensional probability distribution.

7.3. The three-dimensional case

In order to clarify why the mutual information for a three-dimensional PDF can be negative, we will use the example proposed by Matsuda et al. of a three spin system that can be up or down having an interaction among them ruled by a Heisenberg Hamiltonian: $H = -J(x_1x_2 + x_2x_3 + x_1x_3)$, where $J = \pm 1$. As in that case, we will study the case where variables x_i can take only the values ± 1. For this Hamiltonian the shape of the three-dimensional PDF will be very simple since it is formed by eight 3D pixels, called voxels from now on, whose state can be easily calculated using:

$$p(x_1, x_2, x_3) = \frac{e^{-\beta H}}{Z} \tag{12}$$

where $\beta = 1/K_BT$ is related to the inverse of the temperature, and Z is the partition function that, for this case, is $Z = 2e^{3\beta J} + 6e^{-\beta J}$. Since we are interested in keeping things as simple as possible, we calculate the PDF for two extreme cases: when the temperature is infinity ($\beta = 0$) and when the temperature is zero ($\beta = \infty$). For the first case ($\beta = 0$), calculation in both cases $J = \pm 1$ is very simple since all voxels are switched on and therefore the entropy of the systems is maximal and the mutual information is zero.

The interesting case is when the temperature is zero and therefore $\beta = \infty$. In this case we find very different PDF shapes for the cases $J = \pm 1$. We show in panel (a) of figure 7 the PDF for the case $J = 1$. For this case parallel spins are giving the minimum energy of the system and therefore there is a well defined zero Kelvin state with all spins either up or down. In this case the three body mutual information of the system is positive $I_3(X_1, X_2, X_3) = ln2$. The three-dimensional PDF together with its projections in two and one dimensions needed to calculate the mutual information are also shown in panel (a) of figure 7. As it can readily be seen the 3D PDF is relating the variables one by one as in the 2D case, so that the dependency between variables is maximum. Moreover, the projections in two dimensions also allow us to see that there is a relationship between variables, and in fact it fully determines their dependence. Roughly speaking, the 3D PDF is well behaved and its 2D projections are giving us information concerning the dependence between variables.

We will study now the case where $J = -1$ that favors anti-parallel spin interaction: this is a well known case of a geometric frustrated system. In panel (b) of figure 7 we show the three-dimensional PDF for this system when $\beta = \infty$. For this case the mutual information is negative and equal to $I_3(X_1, X_2, X_3) = -0.117$. As it can readily be seen the three-dimensional PDF is not as well behaved as in the last case: if a spin is up, i.e. if we perform a cut on the 3D PDF and take only the upper voxels, there are a lot of possibilities of arranging the other two spins with equal probability. Moreover the projections in two dimensions do not allow us to deduce the shape of the three-dimensional PDF. This causes that the entropies of the two-dimensional projections of the original 3D PDF have a high value and, since they are subtracting terms in the definition of the three variable mutual information of equation 6, the total value is negative.

The main point we would like to emphasise with this very simple example is that a negative mutual information implies a correlation between three variables that can not be seen as the sum of a pairwise dependence of two variables. In other words we must have a careful look at the 3D probability distribution to get any information about how variables are correlated, since its two-dimensional projections have all the information messed up resulting in a high entropic PDF. Fol-

lowing this idea we present in panel (c) of figure 7 a PDF with the maximum negative mutual information: its two-dimensional projections have a maximum entropy and their mutual information is zero, i.e. little about the 3D PDF can be inferred from the 2D projections. However the PDF has a well defined shape in three dimensions, and thus the mutual information is negative and maximal ($I_3(X_1, X_2, X_3) = -\ln 2$).

8. Acknowledgments

This work was supported by the Spanish Ministry of Science and Innovation (Grant FIS2011-24439) and the Catalan Government (Grant 2009SGR-1251).

[1] M. Leocmach, H. Tanaka, Roles of icosahedral and crystal-like order in the hard spheres glass transition, Nat. Commun. 3 (2012) 974.
[2] H. Tanaka, Importance of many-body orientational correlations in the physical description of liquids, Faraday Discuss 167 (2013) 9-76.
[3] J.P. Hansen, I.R. McDonald, Theory of simple liquids, Academic Press, London (1986).
[4] A.K. Soper, An asymmetric model for water structure, J. Phys.: Cond. Mat. 17 (2005) S3273.
[5] M. Leetmaa, K.T. Wikfeldt, M.P. Ljungberg, M. Odelius, J. Swenson, A. Nilsson, L.G.M. Pettersson, Diffraction and IR/Raman Data do not Prove Tetrahedral Water, J. Chem. Phys. 129 (2008) 084502.
[6] K.T. Wikfeldt, M. Leetmaa, M.P. Ljungberg, A. Nilsson, L.G.M. Pettersson, On the Range of Water Structure Models Compatible with X-ray and Neutron Diffraction Data, J. Phys. Chem. B 113 (2009) 6246.
[7] A.K. Soper, Orientational correlation function fot molecular liquids: the case of liquid water, J. Chem. Phys. 101 (1994) 6888.
[8] A.K. Soper, M.A. Ricci, Structures of high-density and low-density water, Phys. Rev. Lett. 84 (2000) 2881.
[9] Ph. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L.A. Naslund, T.K. Hirsch, L. Ojamae, P. Glatzel, L. G. M. Pettersson, A. Nilsson, The Structure of the First Coordination Shell in Liquid Water, Science, 304 (2004) 995-999.
[10] M.F. Chaplin, A proposal for the structuring of water, Biophys. Chem. 83 (1999) 211.
[11] I.M. Svishchev, H. Kusalik, Structure in liquid water - a study of spatial-distribution functions, J. Chem. Phys. 99 (1993) 3049.
[12] A. De Santis, D. Rocca, The local order in liquid water studied through restricted averages of the angular correlation function, J. Chem Phys. 107 (1997) 9559.
[13] A. De Santis, D. Rocca, Angular distribution functions and specific local structures in liquid water, J. Chem Phys. 107 (1997) 10096.
[14] T. Lazaridis, M. Karplus, Orientational correlations and entropy in liquid water, J. Chem. Phys. 105 (1996) 4294.
[15] M. Rovira-Esteva, L.C. Pardo, J.Ll. Tamarit, N. Veglio, F.J. Bermejo, Metastable systems under pressure 63 (2010).
[16] A.H. Narten, M.D. Danford, H.A. Levy, Structure and Intermolecular Potential of Liquid Carbon Tetrachloride Derived from X-Ray Diffraction Data, J. Chem. Phys. 46 (1967) 4875.
[17] P.A. Egelstaff, D.I. Page, J.G. Powles, Orientational correlations in molecular liquids by neutron scattering carbon tetrachloride and germanium tetrabromide, Mol. Phys. 20 (1971) 881.
[18] P. Jedlovszky, Structural study of liquid methylene chloride with reverse Monte Carlo simulation, J. Chem. Phys. 107 (1997) 7433.
[19] P. Jóvári, G. Mészáros, L. Pusztai, E. Sváb, The structure of liquid tetrachlorides CCl4, SiCl4, GeCl4, TiCl4, VCl4, and SnCl4, J. Chem. Phys. 114 (2001) 8082.
[20] L.C. Pardo, N. Veglio, F.J. Bermejo, J.Ll. Tamarit, G.J. Cuello, Experimental assessment of the extent of orientational short-range order in liquids, Phys. Rev. B 72 (2005) 014206.
[21] R. Rey, Quantitative characterization of orientational order in liquid carbon tetrachloride, J. Chem. Phys. 126 (2007) 164506.
[22] Sz. Pothoczki, L. Temleitner, P. Jóvári, S. Kohara, L. Pusztai, Nanometer range correlations between molecular orientations in liquids of molecules with perfect tetrahedral shape: CCl4, SiCl4, GeCl4 and SnCl4, J. Chem. Phys. 130 (2009) 064503.
[23] G. Evrard, L. Pusztai, Reverse Monte Carlo modelling of the structure of disordered materials with RMC++: a new implementation of the algorithm in C++, J. Phys.: Condens. Matter 17 (2005) S1; O. Gereben, P. Jóvári, L. Temleitner, L. Pusztai, A new version of the RMC++ Reverse Monte Carlo programme, aimed at investigating the structure of covalent glasses, J. Optoel. Adv. Mater. 9 (2007) 3021; L. Pusztai, R.L. McGreevy, On the structure of simple liquids SbCl5 and WCl6, J. Chem Phys. 125 (2006) 044508; L. Temleitner, L. Pusztai, Local order and orientational correlations in liquid and crystalline phases of carbon tetrabromide from neutron powder diffraction measurement, Phys. Rev. B 81 (2010) 134101.
[24] A.K. Soper, Partial structure factors from disordered materials diffraction data: An approach using empirical potential structure refinement, Phys. Rev. B 72 (2005) 104204.
[25] B. Hess, C. Kutzner, D. Van der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, J. Chem. Theory Comput 4 (2008) 435.
[26] S. Busch, L.C. Pardo, W.B. O'Dell, C.D. Bruce, C.D. Lorenz, S.E. McLain, On the structure of water and chloride ion interactions with a peptide backbone in solution, Phys.Chem. Chem. Phys., 5 (2013) 21023.
[27] M. Rovira-Esteva, N.A. Murugan, L.C. Pardo, S. Busch, J.Ll. Tamarit, Sz. Pothoczki, G.J. Cuello, F.J. Bermejo, Interplay between intramolecular and intermolecular structures of 1,1,2,2-tetrachloro-1,2-difluoroethane, Phys. Rev. B 84 (2011) 064202.
[28] M. Rovira-Esteva, N.A. Murugan, L.C. Pardo, S. Busch, M.D. Ruiz-Martn, M.S. Appavou, J.Ll. Tamarit, C. Smuda, T. Unruh, F.J. Bermejo, G.J. Cuello, S.J. Rzoska, Microscopic structure and dynamics of high and low density trans-1,2-dichloroethylene liquids, Phys. Rev. B 81 (2010) 092202.
[29] C.E. Shannon, A Mathematical Theory of Communication. Bell System Technical Journal 27 (1948) 379423.
[30] C.E. Shannon, Communication Theory of Secrecy Systems. Bell System Technical Journal 28 (1949) 656715.
[31] P. Haggerty, The Corporation and Innovation, Strategic Management Journal, 2 (1981) 97-118.
[32] J.L. Kelly, Jr., A New Interpretation of Information Rate, Bell System Technical Journal, 35 (1956) 917-26.
[33] H. Matsuda, Physical nature of higher-order mutual information: Intrinsic correlations and frustration, Phys. Rev. E, 62 (2000) 3096.
[34] P.G. Debenedetti, F.H. Stillinger, Supercooled liquids and the glass transition, Nature, 410 (2001) 259-267.

[35] F.Sausset, G. Tarjus, P. Viot, Tuning the Fragility of a Glass-Forming Liquid by Curving Space, Phys. Rev. Lett., 101 (2008) 155701.